Research In Intelligent System: Problems In Gathering Knowledge

Fadzilah Siraj[a], Wan Hussain Wan Ishak[b], Fadhilah Mat Yamin, Zulhazlin Abas
School of Information Technology,
Universiti Utara Malaysia, 06010 Sintok, Kedah, MALAYSIA
Email: [a]fad173@webmail.uum.edu.my, [b]hussain@uum.edu.my

Abstract

Artificial intelligence or AI has come to its era of popularity since its introduction in 1930s. The era of AI started as the theoretical concept, later implemented in various applications. Initial research in AI concentrates on emulating human ability and thinking. In more advanced study of AI, the potential of human "technology" were further exploited. However, as human "technology" uses a large amount of knowledge (gathered from one's experience), intelligent system could fail if knowledge provided to the system is insufficient. This paper discusses the important of knowledge in intelligent system development. Based on our research in developing intelligent systems, problems in gathering knowledge are addressed. Solutions for future research in this area are proposed.

1.0 Introduction

Artificial Intelligence or AI is one of the greatest founding in computer science. A number of computer science fields and technologies have been used in developing intelligent systems, starting from traditional information systems and databases, to modern distributed systems and Internet. In AI, the technology is developed based on the combination of knowledge from various disciplines such as biology, physics, mathematics and statistics. Artificial Neural network for example, was established by a combination of knowledge from statistics, mathematics, biology and physics. Therefore, knowledge is a very important component in AI. It resides in many different places such as: databases, knowledge bases, filing cabinets and peoples' heads.

In early 1940s, researchers started to review the potentials of human technology and the ability of computer programs to act intelligently. In 1950s Alan Turing presents an experiment called Turing Test to demonstrate the ability of machine acting like human. The experiments encouraged many other researchers to investigate the potentials of human technology in machine. However, human technology runs based on knowledge. Human knowledge on the other hand is gathered from one's experience or from written knowledge (written documents such as books, magazine, news paper and etc.). Written knowledge is written by expert or someone with experience or knowledge in the targeted field. Most of the written knowledge is highly structured where documents are usually well organized and indexed. Written knowledge is very useful but in some cases it requires further explanation or illustration, which can be done by human. Human knowledge is a very complex knowledge that is some kind of tacit knowledge where people accumulate as they do their jobs.

Knowledge gathering is one of the ways to obtain expert knowledge. Its aims is to make tacit knowledge explicit and is well suited to eliciting information for a clearly define purpose. The ultimate goal of knowledge gathering is to capture an individual's decision-making process with enough clarity that someone else guided by it could repeat the steps of the process and achieve the same result (Eisenhart, 2001). This paper discusses the importance of knowledge in intelligent system development. Based on our research in developing intelligent system, problems in gathering knowledge are addressed. Solutions for future research in this area are also proposed.

2.0 Knowledge and Requirements

Knowledge base constitutes of different kinds of knowledge, namely Control Knowledge (C Knowledge), Domain Knowledge (D Knowledge), Explanatory Knowledge (E Knowledge) and System Knowledge (S Knowledge). C Knowledge refers to system's problem solving strategies and its functional model. D Knowledge refers to application domain facts, theories and heuristics. E knowledge defines the contents of explanations and justifications of the system's reasoning process, as well as the way they are generated. S Knowledge describes the contents and structure of the knowledge base. In some intelligent systems, system knowledge also defines user models and strategies for the system's communication with its users.

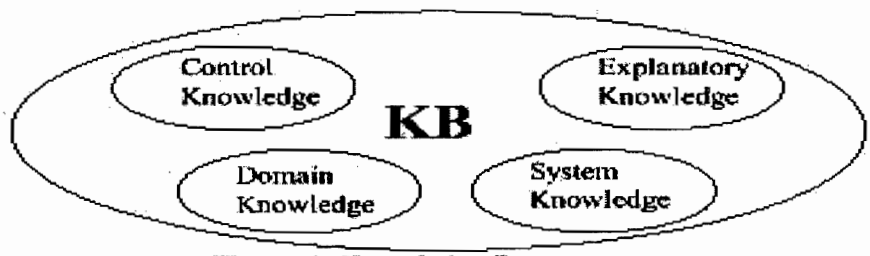

Figure 1: Knowledge Base contents

The development of knowledge processing in AI is governed by the development of the knowledge processors and the integration of traditional programming languages with knowledge-processing capabilities. The rest of this paper discusses several ongoing researches at School of Information Technology, Northern University of Malaysia.

2.1 Intelligent Tutoring System

Intelligent Tutoring Systems (ITS) is one of the AI techniques that has been developed as an alternative to Computer-Assisted Instruction (CAI) (Ibrahim *et al.*, 1989; Meenakumari and Radhakrishna, 1991). ITS exhibits intelligent in the following ways (Meenakumari and Radhakrishna, 1991):

- ITS has knowledge of the subject being thought (Domains Expertise).
- ITS has current information regarding the student knowledge (Student Model).

- ITS can articulate the solution or the thinking process and can communicate with the students (Communication Expertise).
- ITS possesses the knowledge regarding the pedagogical aspects of teaching and be able to choose the best teaching strategy (Teaching Knowledge Expertise).
- ITS has learning capability (Learning Expertise)
- ITS has the heuristic problem solving capability.

ITS has the capability to diagnose difficulties, misconceptions or lack of understanding and provide appropriate remediation (Kit *et al.*, 1991). It consists of Expert Systems module, pedagogy module, student model and user interface (see Figure 2). Expert System is the main component in ITS (Clancey, 1981 dan Clancey, 1986). It has the ability to guide the student in solving the problems and measure the student's performance. Pedagogy module is used to control the interaction between the student and the system. Student model is used to determine students' levels of understanding with regard to the subject being thought and their progress during the teaching and learning process. Communication between student and systems is through the interface. ITS will provide the students with a suitable course materials and guides the students to increase their performance (see Warendorf and Tan, 1997; Stern and Woolf, 1998). The utilization of ITS for teaching and learning could also reduce instructor workloads (Koedinger *et al.*, 1997).

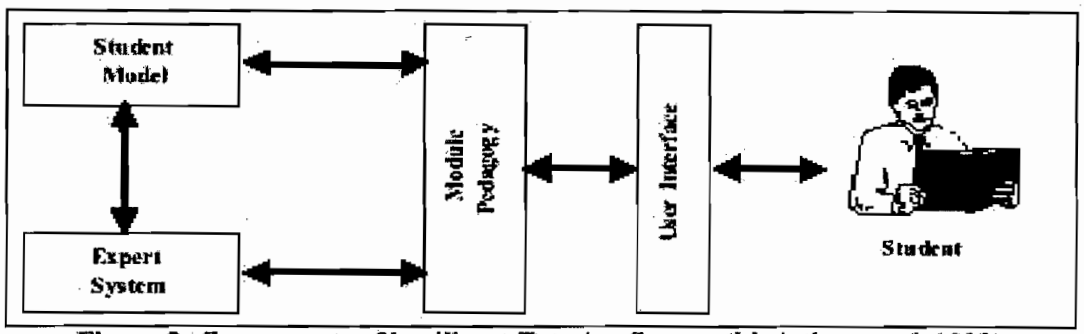

Figure 2: Components of Intelligent Tutoring System (McArthur *et. al.*, 1993)

Tutor (sometimes is referred as instructor) in context of student and tutor interaction is referring to the "tutor" and human tutor. Student will have direct communication with the "tutor", however, there might be some problems that the ITS cannot solve. The performance of ITS mainly depends on the knowledge; knowledge about the student background, pedagogical knowledge and expert knowledge. At this stage human tutor will have to play his part. The unsolved problem will show the current weakness of the "tutor" reasoning module and the human tutor could take the appropriate action to update the "tutor" knowledge. In this case, human tutor or the organization need to work together with system developer to make sure that ITS provides the best performance at all times.

2.2 Natural Language Processing

Natural language processing is a process that can make computer understands natural language. It can be divided into two parts namely natural language understanding and

natural language generation (Cawsey, 1997). The scope of natural language understanding involves syntax of spoken or written sentences, as well as its semantic.

According to Allen (1995) there are seven different forms of knowledge relevant for natural language understanding, viz;

- Phonetic and phonological knowledge – concerns how words are related to the sounds
- Morphological knowledge – concerns how words are constructed from basic meaning units called morphemes. A morpheme is a primitive unit of meaning in a language.
- Syntactic knowledge – concerns how words can be put together to form correct sentences and determines what structural role each word plays in the sentence and what phrases are subparts of what other phrases.
- Semantic knowledge – concerns what words mean and how these meanings combine in a sentence to form sentence meanings.
- Pragmatic knowledge – concerns how sentence is used in different situations and how use its meaning affects the interpretation of the sentence.
- Discourse knowledge – concerns how the immediately preceding sentence affects the interpretation of the next sentence.
- World knowledge – includes the general knowledge about the structure of the world.

Phonetic and phonological knowledge are very important in developing speech recognition system. Using this knowledge, the system can identify what are the users want to say by the sound that the user's made. To identify this, first the system gets the sound from the user in analogical signal. This signal will be converted to digital signal before the system could process the signal. Without this knowledge it will be impossible for the system to know what the sounds are.

The rest of the knowledge is important because to understand natural language, we will need to know the roots of the words. This is achieved by having the morphological knowledge. The syntactic knowledge will then look at the collection of words that formed a sentence. It also looks at the grammar in order to determine the structure of the sentence. The meaning of the sentence will be acquired from semantic knowledge. Hence, the semantic knowledge alone cannot determine the true meaning of the sentence. In natural language, there are sentences that give different meanings in different situations. For example the utterance "Do you know what time it is?" could be a criticism or a genuine question. When facing this problem, pragmatic knowledge is very important to handle it.

Natural Language processing has been proposed as one of the component in intelligent interface in Fadzilah and Wan Hussein (2000) and Fadzilah *et al.* (2000a). For more discussion on the intelligent interface the readers could refer to Fadzilah *et al.* (2000b). In Fadzilah *et al.* (2000c), we discuss the model of natural language processing as an interface to ITS to support distance learning. From the exploratory study, we conclude that natural language processing is one of the most potential techniques in conjunction with the current input processing technique. Figure 3 illustrates the importance of knowledge in natural language processing as an interface for ITS.

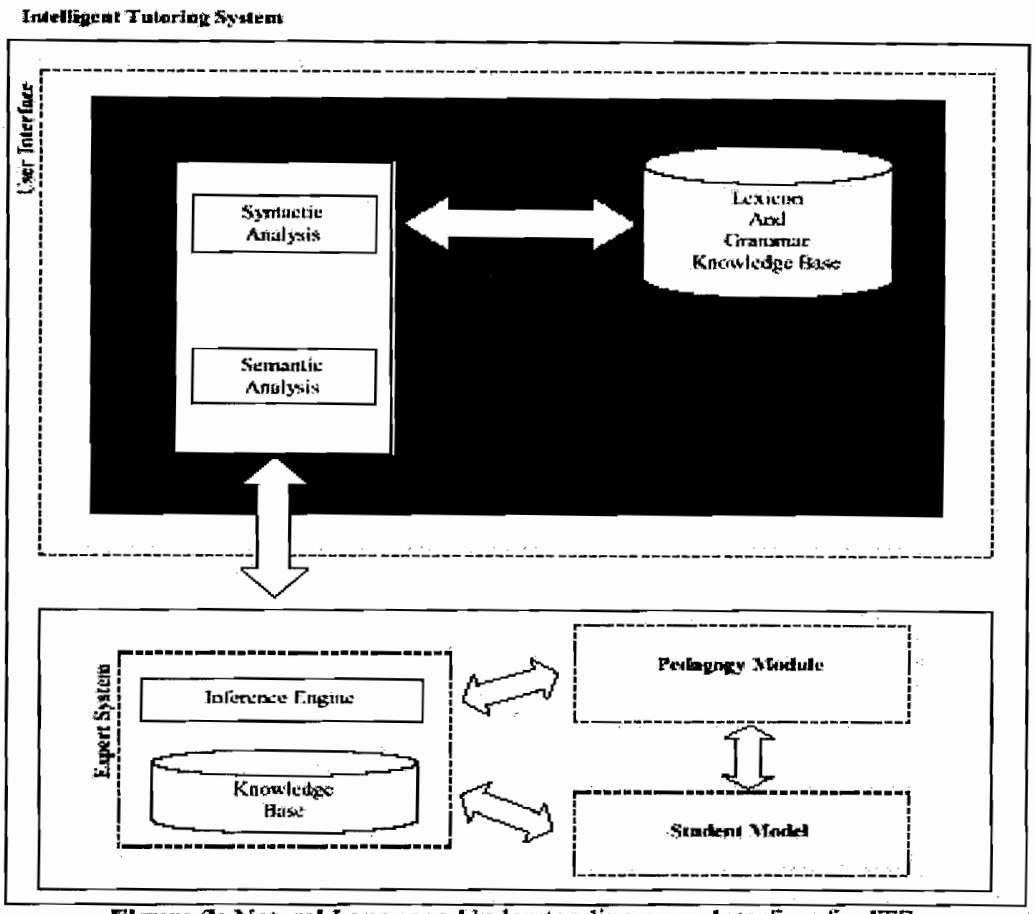

Figure 3: Natural Language Understanding as an Interface for ITS

Another applications of natural language processing is in information abstraction. The application reads through the whole article or story and summaries the article. It is important to identify whether the sentences in the article are related to each other. Hence, the discourse knowledge will be used to handle these problems. In addition, the world knowledge is very important to understand the true meaning of natural language. The bigger the world knowledge is, the better the system will be. However the bigger knowledge means the bigger space is require to store them.

2.3 Information Retrieval

Information Retrieval (IR) plays an increasingly in applications that range from retrieval of textual information from databases, to navigation in a network of information sources such as Internet. IR is a branch of computer science that searches for relevant documents from document databases (Keim *et al.*, 1997). The purpose of information retrieval system is to identify published documents that relevance to the user's needs. Accordingly, an intelligent information retrieval system must be able to

represent user's need and be able to estimate the relevance of a particular document. In a typical IR system, the user expresses the needs by entering the query, either in free text or in some specialized query language. The system attempts to satisfy the query by finding documents that are similar or match the query.

The IR project has three fundamental components: data gathering, system development and evaluation. Data gathering is the most importance stage in IR. Data can be gathered from any resources such of literature review, observation, instruments or data warehouse. Collected data will become knowledge. Knowledge is importance in IR research. Knowledge has their own function according to how they will be implemented. Florescu *et al.* (1998) reveal three classes of task related to information management on the WWW such as modeling and querying the web, information extraction and integration, and web site construction and restructuring. In modeling and querying the web stage, the web acts as a directed graph whose nodes are web pages and edges are links between pages. The queries can be based on the content of the desired page and on the link structure connecting the pages. Generally, a database concepts and technology can be applied in building, restructuring and manage web sites at a final stage.

Catarci *et al.* (1998) develop Web - At - a - Glance (WAG) which exploit a suitable integration of ideas and techniques coming from both the database and the knowledge representation areas. WAG is based on three key issues namely, the building of conceptual representation for any domain of interest, which is populated with data extracted from the web, the availability of a visual interface to easily query such a materialized view and the adoption of both an internal modeling of the information and several acquisition mechanisms based on Description Logic.

Our research in IR focuses on feedback agent that utilize neural network learning capability (Figure 4). The knowledge for the agent is a collection of documents in document database. Extracting the knowledge form the database involves several steps; attributes identification, knowledge extraction and knowledge representation.

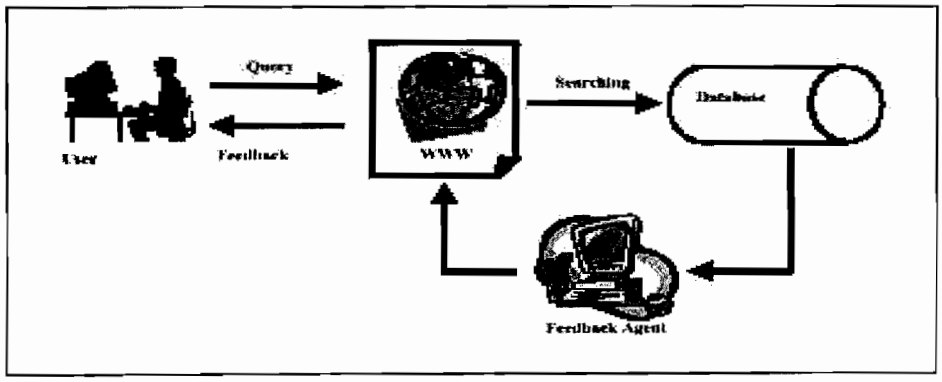

Figure 4: General Architecture Intelligent Information Retrieval

2.4 Medical Applications

To date, most of the medical systems developed were standalone applications with specific databases for certain diseases. This implies that patients information in one system can only be used by that particular system. On the other hand, other systems require another databases for other patients or for the same patients whose records were kept in other databases. Another problem with standalone database is that, the database for the same system in another places would differ as the number of patients using the systems increases. This problem affects the knowledge acquires from the databases and thus influences the decision made by the system. When the number of patients is high, indirectly the system will produce more accurate results compared to the system with less number of patients.

The patients records are valuable information for the knowledge-based system. In fact, the current patients data would enhance and strengthen the validity of the system reasoning (Maniekam and Abidi, 1999). For example, health-care providers could access the electronic record and the data could be stored and updated frequently. By using this method, the system knowledge will always be updated. In intelligent medical system, knowledge gathering is a very challenging task. As the system is highly depended on the knowledge, insufficient knowledge would cause the system fail to perform in specific tasks.

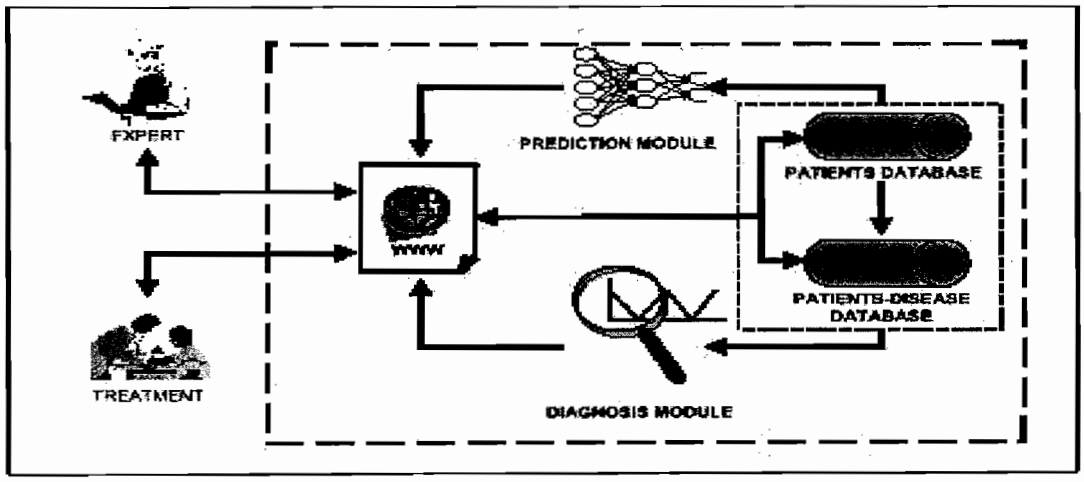

Figure 5: A model for Web-Based Medical Diagnosis and Prediction

The proposed model for Web-Based medical diagnosis and prediction (*see* Figure 5) consists of four components, they are databases, prediction module, diagnosis module and user interface. The databases consist of patients database and patients-disease database. Patients database will be used to store patient's information such as name, addresses, and others particulars details. Patients-disease database stored all information about patients and their illness. The information stored in the database includes types of diseases, the treatments and other details about the test and administering therapy. Patients information are separated in a different database to enhance the patients records storage, so that other departments could use the records when the patients are referred to them. This method could prevent other departments

or unauthorized users from accessing the information about patients diseases and provide a centralized information access for the patients records.

Prediction module utilizes neural networks techniques to predict patients illness or conditions based on the previous similar cases. Data from the patients and patients-disease database will be used for training and testing. The weight from the training will be stored to predict a new data fed into the system. Diagnosis module consists of expert system and fuzzy logic techniques to perform diagnosis tasks. A set of rules will be defined using the patients and patients-disease databases as well as the expert knowledge on the disease domain. Expert system uses the rules to diagnose patient's illness based on their current conditions or symptoms. In addition, fuzzy logic is integrated to enhance the reasoning when dealing with fuzzy data. The combination of expert system and fuzzy logic that forms a hybrid (expert-fuzzy) system could increase the system performance.

3.0 Problems and Solutions

Typically, development of intelligent system and research involved two aspects: knowledge acquisition and knowledge representation (Bielawski and Lewand, 1991). Knowledge acquisition is a process of gaining knowledge with collaboration between knowledge engineer and domain expert. There are two main sources of knowledge: experience and documentation. Experience knowledge (tacit knowledge) is the most popular kind of knowledge being captured in intelligent system development. However, AI techniques such as Neural Networks, Expert System, Fuzzy Logic, Genetic Algorithm, Natural Language Processing and Intelligent Tutoring System require explicit knowledge rather than tacit knowledge. Explicit knowledge is the knowledge that is written down or expressed in some tangible forms (Skyrme, 1997). However, as most of the knowledge is in a form of tacit knowledge, extracting such knowledge is very difficult (Eisenhart, 2001). Knowledge representation is another issue in intelligent system. According to Woods (1986), the representation of knowledge is crucial due to the primitive representation of knowledge, limitation of computer's ability to perceive, know or understand.

Knowledge modeling is another popular process in intelligent system that has received most attention in recent years among developers of intelligent systems, AI practitioners and researchers. Any discussion of knowledge modeling should also include a reference to the kinds of knowledge that can be represented in the knowledge base of an intelligent system, as well as to the basic representation and design techniques.

It is also difficult and time consuming to construct new knowledge bases from scratch. According to Cooke and McDonald (1990) the problems are not because of time-consuming and tedious processes but the weak of knowledge acquisition methods typically used. On the other hand, it is usually difficult to share knowledge encoded in knowledge bases among different knowledge-based systems. The problems arise due to large diversity and heterogeneity of knowledge representation formalisms and the portability of the protocols that provide interoperability between different knowledge based systems and conventional softwares as well as,

mismatching of terminology, taxonomy and higher level modeling of different systems.

To overcome some problems in developing intelligent systems, standard models, taxonomies, vocabularies and domain terminologies must be established in order to develop appropriate knowledge and reasoning modules. These models can be reused for assembling knowledge-based systems instead of building the models from scratch. In other words, there is a need for ontologies or explicit representations of domain concepts as well as, rules for combining terms and relations to define extensions to the vocabulary. Ontologies are particularly useful in collaboration, interoperation, education and modeling.

4.0 Conclusion

There are many problems associated with knowledge management the enterprises have to play important roles such as (Macintosh, 1999);

- to have an enterprise-wide vocabulary to ensure that the knowledge is correctly understood;
- to be able to identify, model and explicitly represent their knowledge;
- to share and re-use their knowledge among differing applications for various types of users; this implies being able to share existing knowledge sources and also future ones;
- to create a culture that encourages knowledge sharing.

Knowledge management in AI system is however, a little bit different with enterprises knowledge management. The development of advanced intelligent systems require support from other fields such as ontologies and knowledge sharing, knowledge processing, intelligent databases, knowledge discovery and data mining, and distributed AI.

There is a notable increase in integration of AI with other disciplines. This is due to the fact that advancement in knowledge modeling and the fact that many research and development communities have realized that AI is not only useful in knowledge-based systems, but in all software systems.

References

Allen, J. (1995) Natural language understanding: Cummings Publishing Company: Redwood City

Bielawski, L., and Lewand, R. (1991). Intelligent Systems Design: Integrating Expert Systems, Hypermedia, and Database Technology. Wiley Pro. Computing: New York.

Briscoe, T (1995) Robust Parsing, Survey of the State of the Art in Human Language Technology: Oregon Graduate Institute.

Catarci, T., Nardi, D., Santucci, G., and Chang, S., K. (1998). WAG: Web-at-a-Glance. *International Journal of Cooperative Information Systems (IJCIS)*, 7(2 & 3), pp:187-214.

Cawsey, A. (1998) The Essence of Artificial Intelligence: Prentice Hall, Edinburgh.

Clancey, W. J. (1981). Methodology for Building An Intelligent Tutoring Systems. *Technical Report No. STAN-CS-81-894*. Stanford University, Stanford

Clancey, W. J. (1986). Intelligent Tutoring Systems: A Tutorial Survey. *Technical Report No. STAN-CS-87-1174*. Stanford University, Stanford

Cole, R. A., Mariani, J. *et al* (1995) Survey of the State of the Art in Human Language Technology: Oregon Graduate Institute.

Cooke, N. M., and McDonald, J. E. (1990). A Formal Methodology for Acquiring and Representing Expert Knowledge. In Peter G. Reath (Ed.), Expert System: A Software Methodology for Modern Applications, pp: 180-192.

Florescu, D., Levy, A., and Mendelzon, A. (1998). Database Techniques for the World Wide Web: A Survey. SIGMOD Record, Vol. 27, No. 3, pp: 59--74.

Ibrahim, J., Ismail, A. S., Yit, A. Z. M., Kadir, A. K. A., Rahman, A. A., and Salim, N (1989). Kepintaran Buatan dan Pendidikan: Konsep dan Isu. *Proceeding SeminarKepintaran Buatan'89: Persediaan Menghadapi Teknologi Komputer Masa Depan*, Universiti Teknologi Malaysia, Kuala Lumpur, pp. 8-19

Japkowicz, N. (2000) Statistical Natural Language Processing, lecture note

Keim, M., Lewis, D. D., and Madigan, D. (1997). Bayesian Information Retrieval: Preliminary Evaluation. *In Preliminary Papers of the Sixth International Workshop Artificial Intelligence and Statistics*, pp: 303,310.

Kit, L. C., Chee, S. W. and Hoe, A. E. (1991). Intelligent Tutoring Systems-Issues in Development and Use. *Proceeding Conference of the South East Asia Regional Computer Confederation (SEARCC'91): Information Technology Challenges Towards 21st Century*, Bali, Indonesia.

Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M. A. (1997). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education*, 8, pp. 30-43.

Kong, T., E., & Adhaileh, M., H.,(1996). A Flexible Example-Based Parser Based On the SSTC: Universiti Sains Malaysia, Penang.

Macintosh, A. (1999). Knowledge Management. http://www.aiai.ed.ac.uk/~alm/kamlnks.html (April 19, 2001).

Manickam, S., and Abidi, S. S. R. (1999). Experienced Based Medical Diagnostics System Over The World Wide Web (WWW), *Proceedings of The First National Conference on Artificial Intelligence Application In Industry*, Kuala Lumpur, pp. 47 - 56.

Meenakumari , S. and Radhakrishna, M. (1991). Intelligent Tutoring Systems-Issues in Development and Use. *Proceeding Conference of the South East Asia Regional Computer Confederation (SEARCC'91): Information Technology Challenges Towards 21st Century*, Bali, Indonesia.

Siraj, F., Abas, Z., Ishak, W. H. W., and Yamin, F. M. (2000c). Penggunaan Teknik Kepintaran Buatan Bagi Menyokong Penddikan Jarak Jauh. Seminar Kebangsaan Pendidikan Jarak Jauh (31 Okt-01 Nov, 2000). Universiti Kebangsaan Malaysia.

Siraj, F., and Ishak, W. H. W. (2000). Electronic Model for Interactive Teaching and Learning in Cyber Space: Towards Virtual University. *Persidangan Ulang tahun ke-30 UKM (5-6 Sept. 2000)*, Universiti Kebangsaan Malaysia

Siraj, F., Ishak, W. H. W., and Siraj, S. (2000a). Electronic Model with Intelligent Interface for Interactive Web-Based Intelligent Tutoring System. *Learning 2000 (Sept. 28, 29, 30, & Oct. 01)*, Virginia.

Siraj, F., Zakaria, A., and Ishak, W. H. W., (2000b). Intelligent Interface to Support Intelligent Tutoring System. *In Proceeding National Conference on Teaching and Learning in Higher Education*, Universiti Utara Malaysia

Skyrme, D. J. (1997). Knowledge Management: Oxymoron or Dynamic Duo?. *Managing Information*, Vol. 4, No. 7, pp. 24-26

Skyrme, D. J. (1998). Fact or Fad? Ten Shifts in Knowledge Management: Beneath the KM fad lie substantive management practices. *Knowledge Management Review*, pp. 6-7

Stern, M. K. and Woolf, B. P. (1998). Curriculum Sequencing in a Web-based Tutor. *In the Proceedings of Intelligent Tutoring Systems.* http://www-aml.cs.umass.edu/publications.html

Warendorf, K. and Tan, C. (1997). ADIS-An Animated data Structure Intelligent Tutoring System or Putting an Interactive Tutor on the WWW. *Proceedings of the workshop "Intelligent Educational Systems on the World Wide Web" 8th World Conference of the AIED Society*, Kobe, Japan.

Woods, W. A. (1990). Important Issues in Knowledge Representation. In Peter G. Reath (Ed.), Expert System: A Software Methodology for Modern Applications, pp: 180-192.

Zaenen, A and Uszkoreit (1995) Language Analysis and Understanding, Survey of the State of the Art in Human Language Technology: Oregon Graduate Institute.